jeudi 22 mars 2007

Suivi des jobs grille

Atelier régulation de la production dans un contexte grille

dapnia
cea
saclay

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Sommaire

- Monitoring au CC
- Outils de suivi des jobs
- Outils de tracabilité
- Actions entreprises
- Problèmes rencontrés
- Questions

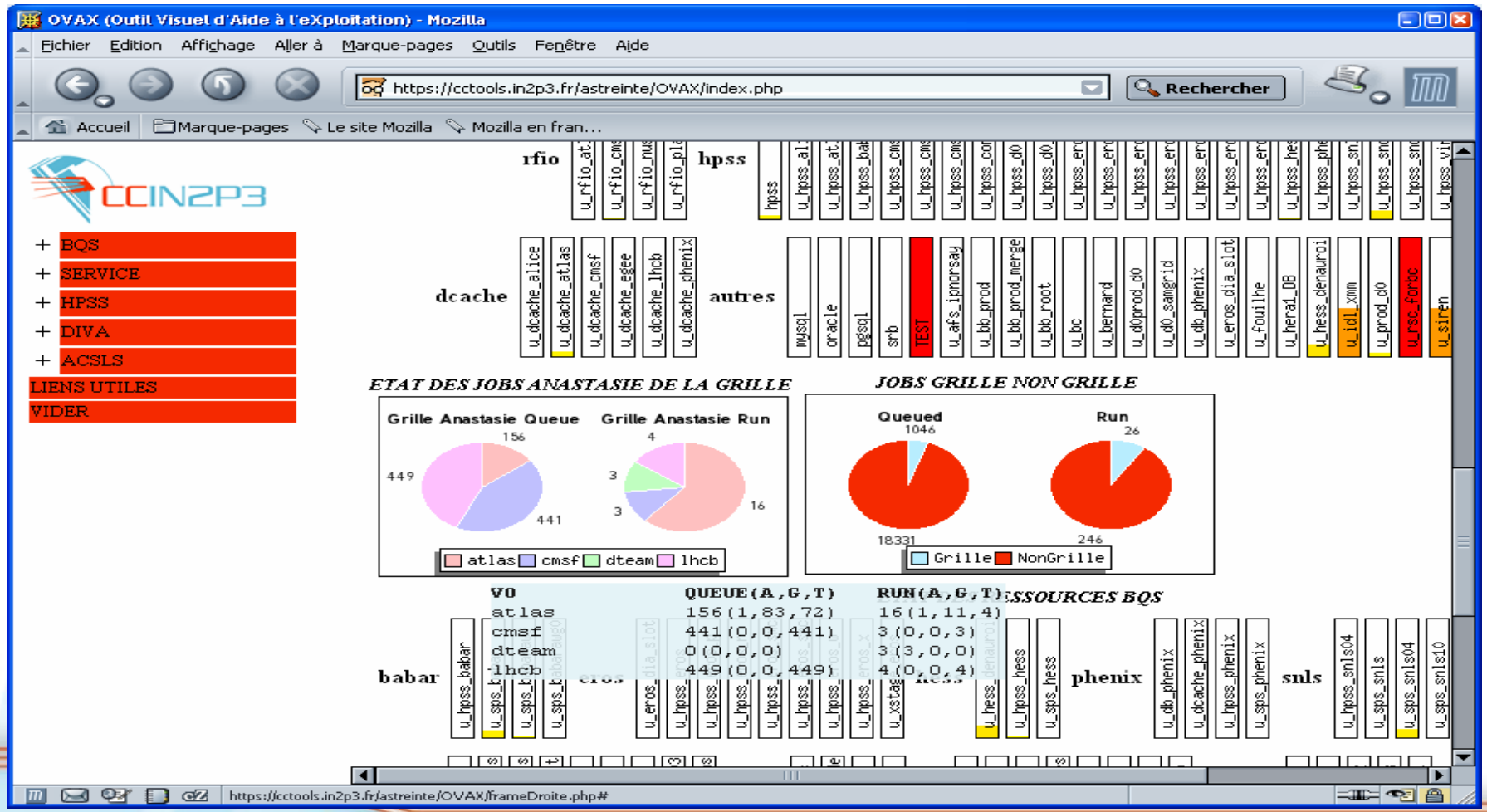
Monitoring au CC

- **État global de la production**

1. Vision graphique globale de l'état des lieux (services, état de la production) : **OVAX**
2. Scripts d'interrogation de l'état du batch
3. Étude statistique de la production en temps réel : **MRTG**
4. État des machines : **SMURF**
5. Outil de *logging* interne au CC : **Web RLS**
6. Outils de monitoring : **NAGIOS** (grille)**, NGOP**
7. Autres outils de monitoring du projet : **Gstat**, CIC *dashboard*, SAM test
8. *Dashboard* **ARDA** décrit l'état du site pour toutes les VOs LHC

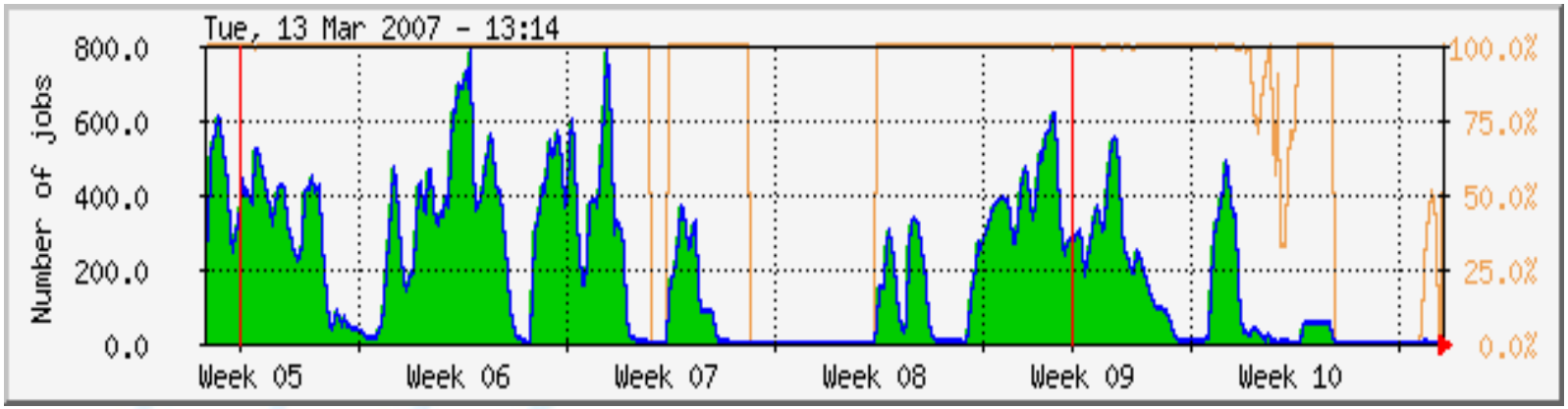
Monitoring au CC : OVAX

- Vision graphique globale de l'état des lieux (services, état de la production)



Monitoring au CC : MRTG

- Étude statistique de la production en temps réel

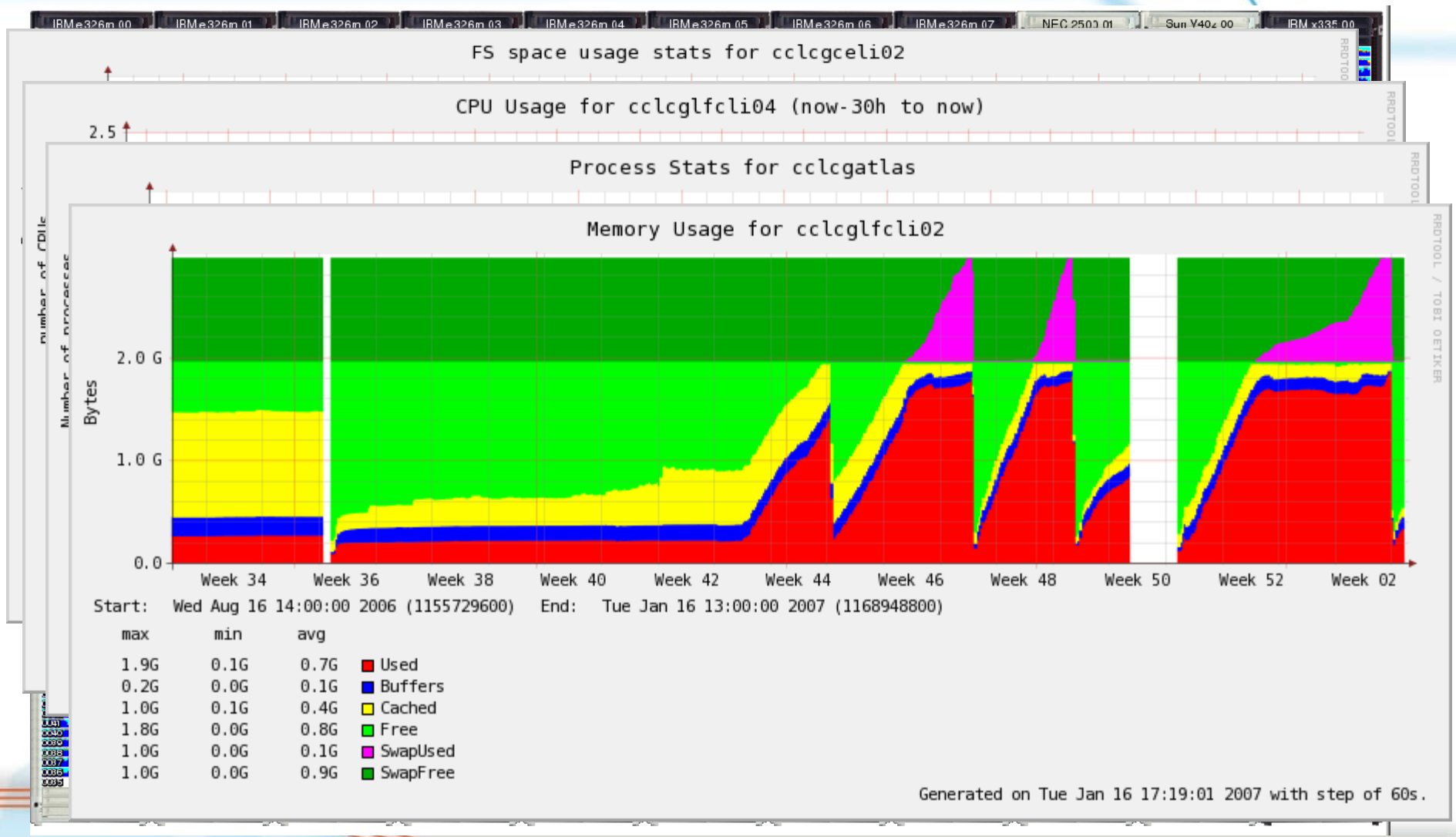


Vert : CMSF grid running jobs at CC-IN2P3

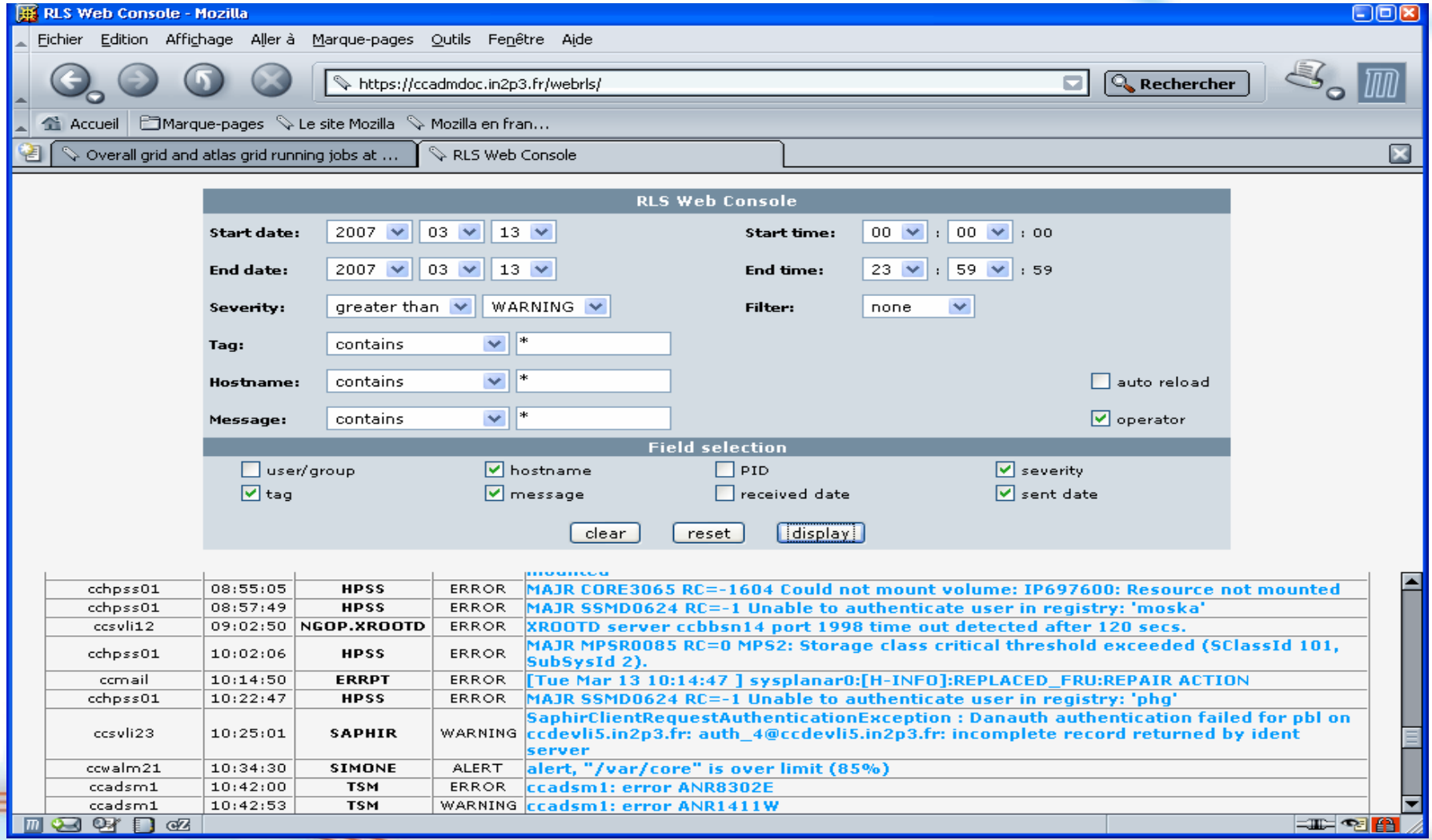
Bleu : All CMSF running jobs at CC-IN2P3

Orange : ((CMSF grid running jobs at CC-IN2P3)/(All CMSF running jobs at CC-IN2P3))*100

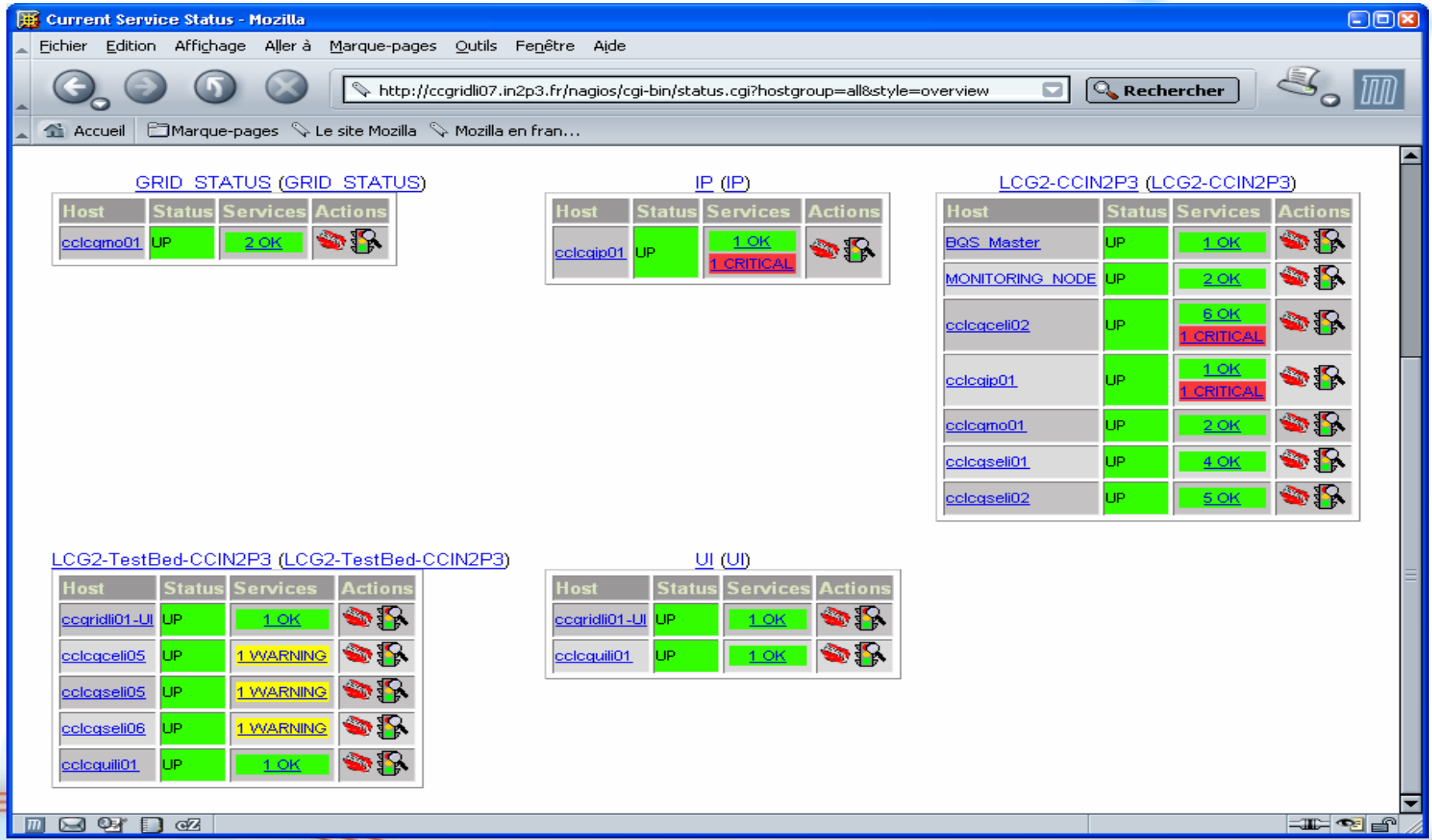
Monitoring au CC : SMURF (état des machines)



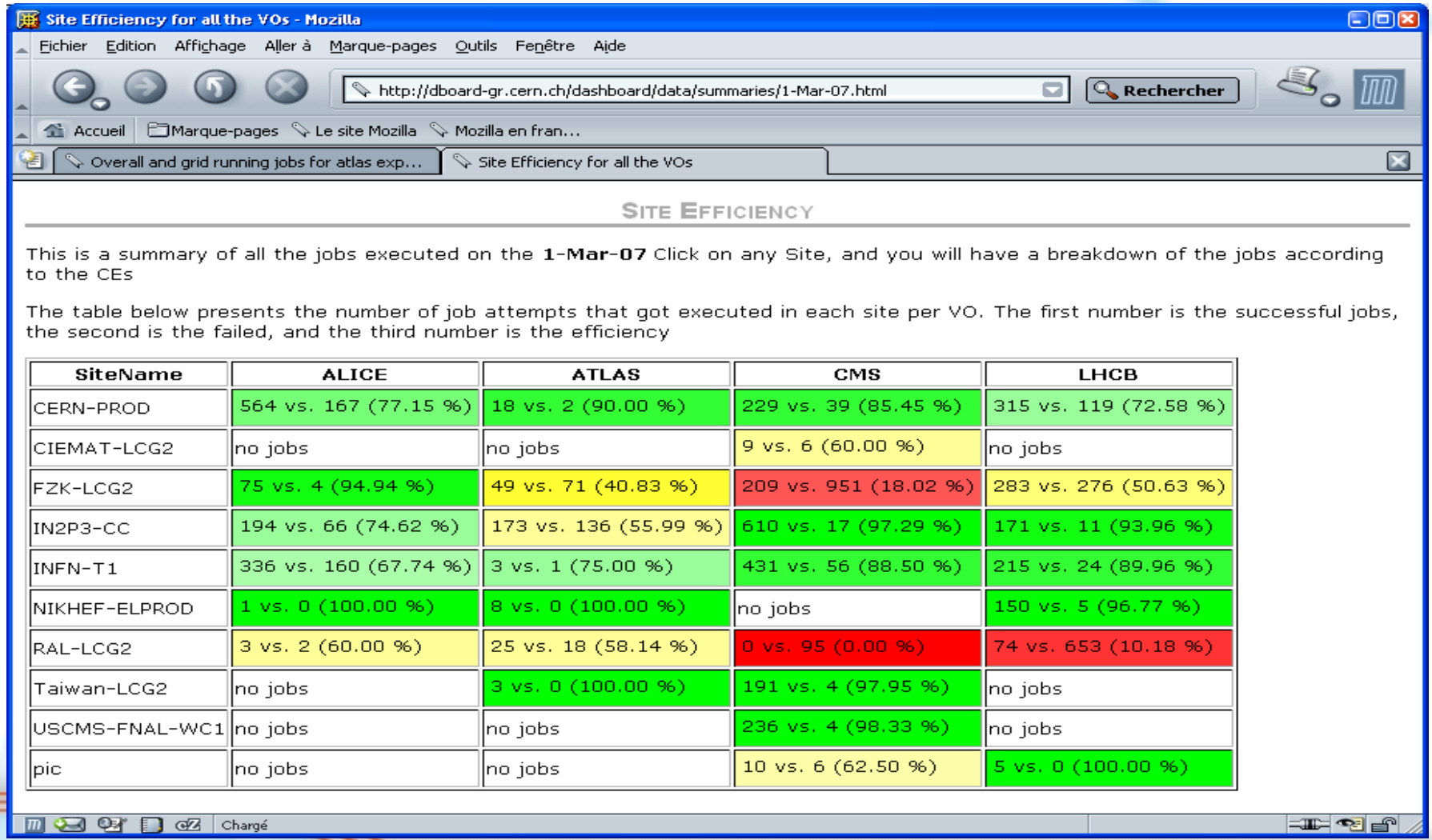
Outil de Monitoring : WebRLS



Monitoring au CC : NAGIOS



Monitoring LHC : ARDA



Site Efficiency

This is a summary of all the jobs executed on the **1-Mar-07** Click on any Site, and you will have a breakdown of the jobs according to the CEs

The table below presents the number of job attempts that got executed in each site per VO. The first number is the successful jobs, the second is the failed, and the third number is the efficiency

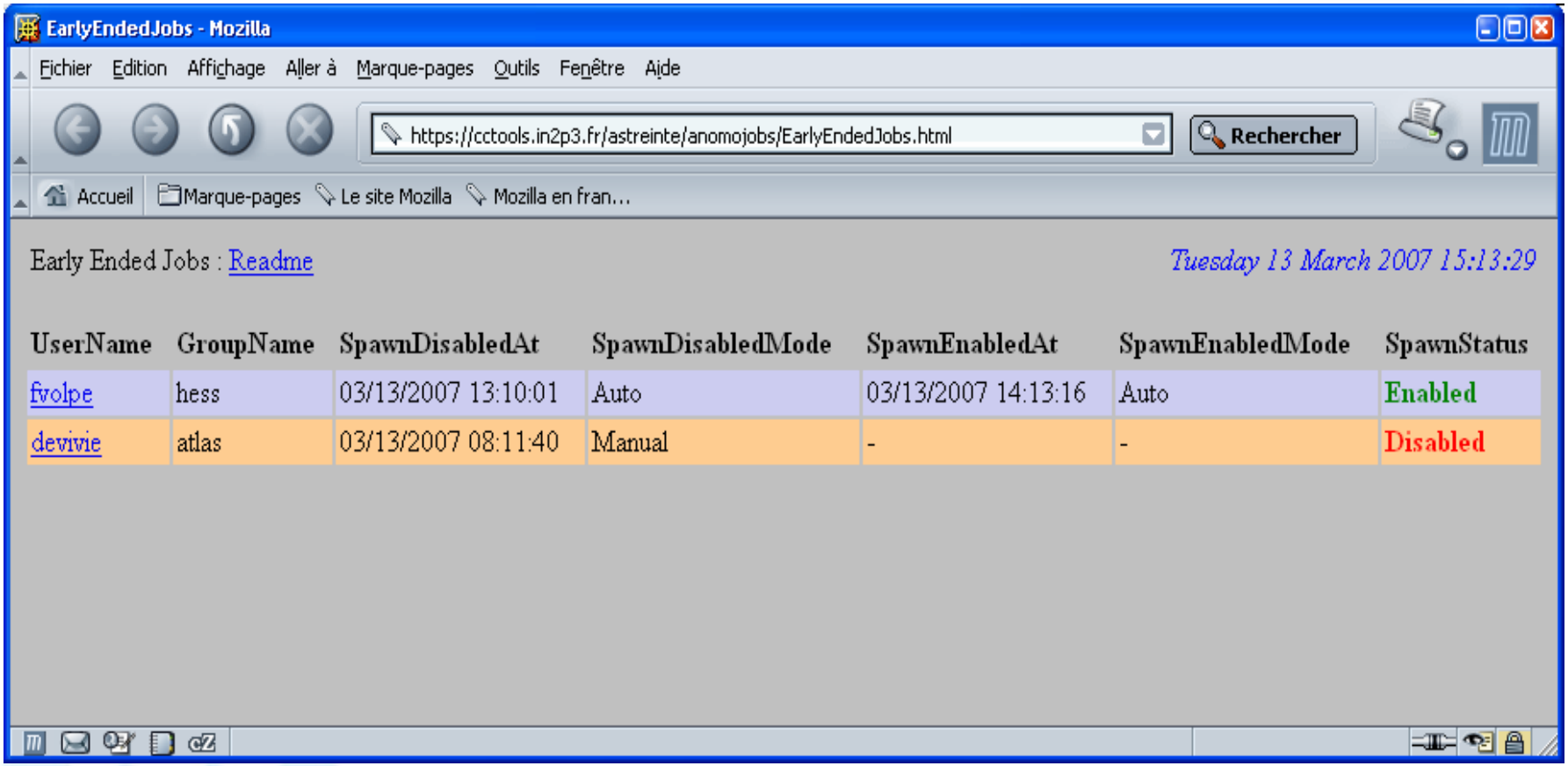
SiteName	ALICE	ATLAS	CMS	LHCB
CERN-PROD	564 vs. 167 (77.15 %)	18 vs. 2 (90.00 %)	229 vs. 39 (85.45 %)	315 vs. 119 (72.58 %)
CIEMAT-LCG2	no jobs	no jobs	9 vs. 6 (60.00 %)	no jobs
FZK-LCG2	75 vs. 4 (94.94 %)	49 vs. 71 (40.83 %)	209 vs. 951 (18.02 %)	283 vs. 276 (50.63 %)
IN2P3-CC	194 vs. 66 (74.62 %)	173 vs. 136 (55.99 %)	610 vs. 17 (97.29 %)	171 vs. 11 (93.96 %)
INFN-T1	336 vs. 160 (67.74 %)	3 vs. 1 (75.00 %)	431 vs. 56 (88.50 %)	215 vs. 24 (89.96 %)
NIKHEF-ELPROD	1 vs. 0 (100.00 %)	8 vs. 0 (100.00 %)	no jobs	150 vs. 5 (96.77 %)
RAL-LCG2	3 vs. 2 (60.00 %)	25 vs. 18 (58.14 %)	0 vs. 95 (0.00 %)	74 vs. 653 (10.18 %)
Taiwan-LCG2	no jobs	3 vs. 0 (100.00 %)	191 vs. 4 (97.95 %)	no jobs
USCMS-FNAL-WC1	no jobs	no jobs	236 vs. 4 (98.33 %)	no jobs
pic	no jobs	no jobs	10 vs. 6 (62.50 %)	5 vs. 0 (100.00 %)

Outils de suivi des jobs

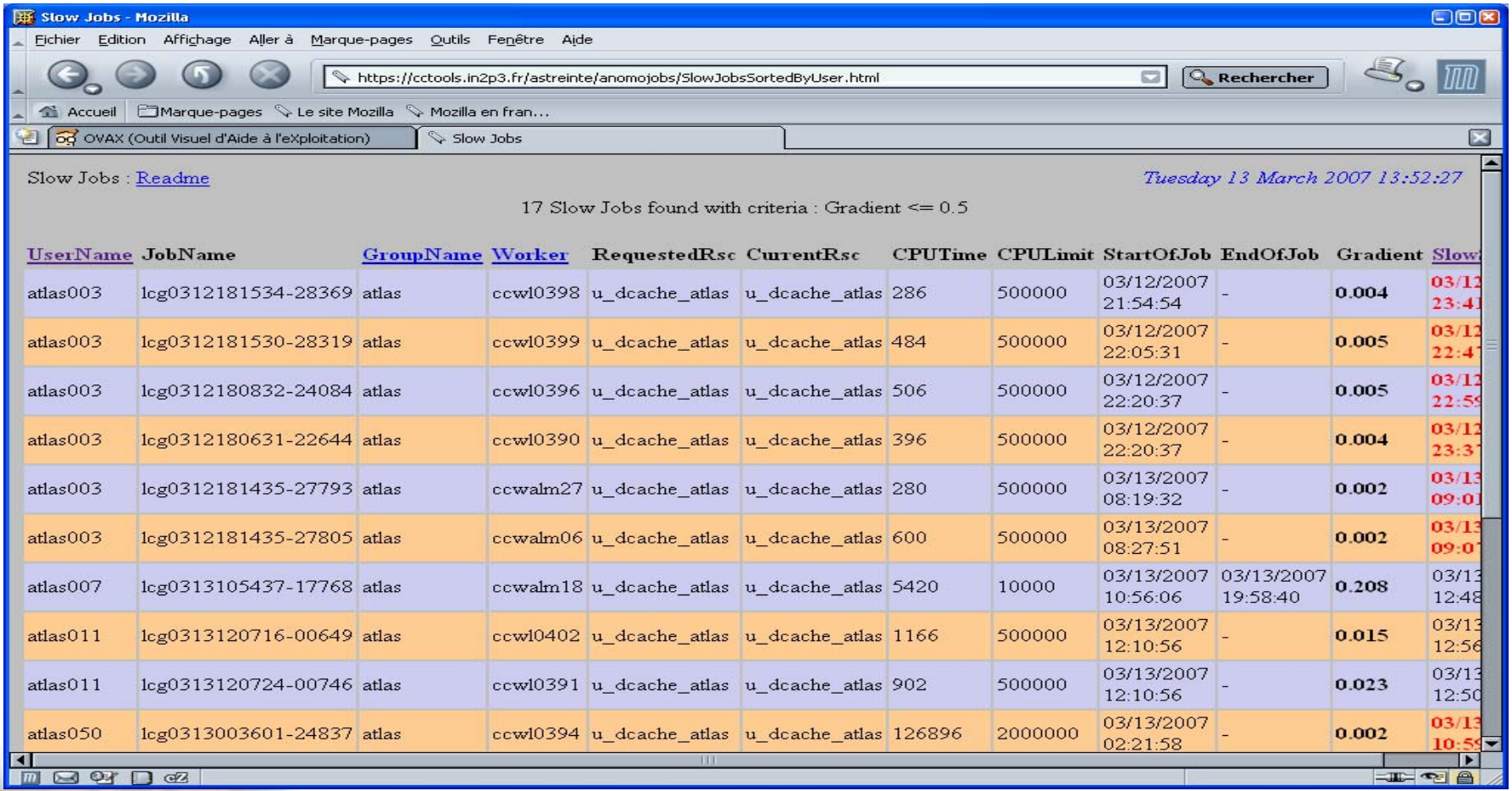
Outils de détection des jobs problématiques

1. **Jobs « slow »** : rapport entre la consommation CPU et le temps de résidence en machine d'un job
2. **Jobs « early ended »** : blocage de l'utilisateur si le nombre de jobs qui ont une consommation CPU très faible est important
   ⇒ problème très probable
3. **Alertes mails :** l'exploitation grille est informée des jobs qui se terminent mal – l'information est envoyée par BQS
4. **Alertes utilisateurs** : sous forme de mail ou de ticket
5. **Alertes sur le statut du site IN2P3-CC** : NAGIOS

Outil de suivi BQS : early ended jobs



Outil de suivi BQS : jobs slow



Slow Jobs : Readme

Tuesday 13 March 2007 13:52:27

17 Slow Jobs found with criteria : Gradient <= 0.5

UserName	JobName	GroupName	Worker	RequestedRsc	CurrentRsc	CPUTime	CPULimit	StartOfJob	EndOfJob	Gradient	Slow
atlas003	lcg0312181534-28369	atlas	ccwl0398	u_dcache_atlas	u_dcache_atlas	286	500000	03/12/2007 21:54:54	-	0.004	03/12 23:41
atlas003	lcg0312181530-28319	atlas	ccwl0399	u_dcache_atlas	u_dcache_atlas	484	500000	03/12/2007 22:05:31	-	0.005	03/12 22:47
atlas003	lcg0312180832-24084	atlas	ccwl0396	u_dcache_atlas	u_dcache_atlas	506	500000	03/12/2007 22:20:37	-	0.005	03/12 22:59
atlas003	lcg0312180631-22644	atlas	ccwl0390	u_dcache_atlas	u_dcache_atlas	396	500000	03/12/2007 22:20:37	-	0.004	03/12 23:37
atlas003	lcg0312181435-27793	atlas	ccwalm27	u_dcache_atlas	u_dcache_atlas	280	500000	03/13/2007 08:19:32	-	0.002	03/13 09:01
atlas003	lcg0312181435-27805	atlas	ccwalm06	u_dcache_atlas	u_dcache_atlas	600	500000	03/13/2007 08:27:51	-	0.002	03/13 09:07
atlas007	lcg0313105437-17768	atlas	ccwalm18	u_dcache_atlas	u_dcache_atlas	5420	10000	03/13/2007 10:56:06	03/13/2007 19:58:40	0.208	03/13 12:48
atlas011	lcg0313120716-00649	atlas	ccwl0402	u_dcache_atlas	u_dcache_atlas	1166	500000	03/13/2007 12:10:56	-	0.015	03/13 12:56
atlas011	lcg0313120724-00746	atlas	ccwl0391	u_dcache_atlas	u_dcache_atlas	902	500000	03/13/2007 12:10:56	-	0.023	03/13 12:50
atlas050	lcg0313003601-24837	atlas	ccwl0394	u_dcache_atlas	u_dcache_atlas	126896	2000000	03/13/2007 02:21:58	-	0.002	03/13 10:59

Traçabilité des jobs

- **Récupération d'information auprès de BQS :**
  - par requête : utilisateur, date de soumission, worker, CE de mise en exécution, profil du job, consommations CPU, mémoire, statut du job, ressources demandées,BQS job ID, certificat de l'utilisateur…
  - par consultation des logs : traces des processus en cours, date, scripts mis en exécution
- **Récupération d'information auprés du CE :**
  - stderr/stdout, LCG job ID, globus-job-ID, consultations des traces logs du globus-gatekeeper
- **Récupération d'information sur le Worker Node**
  - processus en cours d'exécution, output/log du job, connexions en cours,
- **Identification de l'utilisateur :** identité, VO, Email

Création de scripts internes pour faciliter l'accès à ce type d'information

Actions entreprises

- **Constituer un diagnostic précis de la cause des échecs :**
  - manque de ressource, *proxy* périmé, transferts bloqués, services LCG indisponibles
  - problème dans l'environnement du job
- **Identification des jobs problématiques :**
  - LCG job IDs, BQS job IDs, globus job IDs
- **Contacter l'utilisateur ou l'administrateur de la VO**
- **Contacter les responsables des services en cause :**
  - mail, ticket GGUS
- **Diverses opérations de gestion internes de la production:**
  - Destruction/blocage en queue de jobs
  - Blocage des utilisateurs en cas de problème

Actions entreprises

- **Ajustement des objectifs des Vos**
  - Pour des demandes ponctuelles (DCs)

- **Création de ressources internes à BQS**
  - Pour pallier à l'indisponibilité de services internes

- **Mécanismes de réajustements des priorités et des ressources**
  - Sur demande de la VO attribution de priorités en fonction des rôles
  - Après confirmation auprès de l'utilisateur réévaluation des ressources nécessaires au job

Actions entreprises

- **Vérification de l'état du site**
  - vérification de l'information publiée par le système d'information
  - environnement du job sur le WN, services du CE
  - vérification de l'état des SEs, des services critiques hébergés (FTS,VOMS,LFC...)

Problèmes rencontrés

- Difficultés pour contacter l'utilisateur : email introuvable
- Parfois manque de réactivité des utilisateurs
- Parfois utilisateurs mal informés
- Place du *stdout/stderr* non normalisée
- Problèmes récurrents sur les récupérations ou copies de fichiers : indisponibilités des serveurs SRM, des catalogues LFC…
- Difficultés pour nommer les jobs non soumis via des Ressources Brokers
- Méconnaissance de l'information connue par l'utilisateur sur l'état de ses jobs

Problèmes rencontrés

- Manque de visibilité sur l'état des services centraux LCG
- Processus orphelins sur les workers nodes
- Jobs pilotes inactifs
- Gestion des priorités au sein d'un même groupe à mettre en place (rôle VOMS)
- Manque d'outil pour la désactivation de certaines queues de production sur le CE en cas de besoin
- Gaspillage de ressources telles que la mémoire pour la classe longue

Problèmes rencontrés

- **Discussion autour des formules de ranking avec certaines VOs**
- **Impossible de différencier des utilisateurs qui soumettent des jobs avec le même profil puisqu'ils sont mappés sur le même compte.**

Commentaires / Questions